



# Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

## **NIDI data advisor final report and guidelines for sustainable data archiving**

Groenewold, W.G.F.; de Valk, H.A.G.; Liefbroer, A.C.

2016

### ***document version***

Publisher's PDF, also known as Version of record

### ***document license***

CC BY-NC-ND

[Link to publication in KNAW Research Portal](#)

### ***citation for published version (APA)***

Groenewold, W. G. F., de Valk, H. A. G., & Liefbroer, A. C. (2016). NIDI data advisor final report and guidelines for sustainable data archiving. NIDI/KNAW/RUG.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[pure@knaw.nl](mailto:pure@knaw.nl)

# **NIDI data advisor final report and guidelines for sustainable data archiving**

George Groenewold, Helga de Valk & Aat Liefbroer

## **Final Report**

NIDI has a major role in data that are collected in both national and international data collection efforts. Notable examples are the Netherlands Kinship Panel study (NKPS, the Generations and Gender Programme (GGP) and the Integration of the European Second Generation (TIES). These data are collected and used at NIDI but also by researchers all over Europe and the world.

In order to make all these data collection efforts and related output more visible the data advisor at NIDI started with making an inventory of all the existing data that were collected at NIDI currently and in the past decade(s). For each of these data collections the responsible person at NIDI was contacted by the data advisor and the way in which the data are stored were mapped. Current data storage and use of the data was discussed and options for better data storage in light of OA&DD were discussed. This inventory showed that although a substantial number of datasets are already available via DANS for example this did not apply for all. The present approach at NIDI is to archive all project related data and meta-data in project-specific directories and subdirectories on the NIDI network computer (G: and Q: drives). Automatic backup procedures ensure that data and information stored on NIDI network computers will never get lost. Thus, NIDI and NIDI staff already apply proper routines to archive project-related data and meta-data. However, there is yet no NIDI portal and system of standard procedures providing the wider (research) community access to project-related NIDI data and meta-data. Therefore, NIDI project data and meta-data, after project completion, must also be stored *externally*, on the computer servers of DANS (Data Archiving and Networked Services) the KNAW institute specialized in the sustainable archiving of research data and meta-data.

The data advisor made an overview of the datasets for which action is needed together with the relevant persons. In addition to this, he provided a standard guideline for the archiving of datasets which is outlined below. This protocol specifies the principle guidelines to be followed in case of projects where NIDI is the lead partner as well as those cases where NIDI is partner for the Dutch data collection. These procedures also clarify who has to undertake what action in order to ensure that these guidelines are actively implemented.

Third, the data advisor made a start with further depositing data from NIDI research at DANS. This was carried out together with the relevant responsible persons for these projects in order to train them in following this procedure in the future. Finally, in order to inform the staff at the institute and raise awareness of the importance of data archiving first of all the guidelines were drafted by the data advisor in consultation with two theme group leaders at NIDI (Liefbroer & de Valk). Furthermore within the coming months a brown bag seminar will be organized by the data advisor that informs the staff about this and to ensure successful further implementation.

## **Guidelines for sustainable data archiving**

NIDI fully support the 2011 KNAW policy regarding sustainable archiving of research data and their release to the wider research community (i.e. open access). This policy is in line with two fundamental characteristics of sound scientific research: *verifiability* and *reproducibility*. These characteristics will help to decide what to archive and what not.

This guideline pertains to NIDI projects producing *primary data* (via data collected in which NIDI is the lead partner or responsible for the Dutch data collection in case of international data collection efforts). These include data in databases that compile information collected by means of survey questionnaires, by recordings and accounts of focus group discussions, and by in-depth personal interviews. In addition, the guideline also applies to the meta-data produced by NIDI projects, such as questionnaires, code books, documents describing methodological issues such as sample design and implementation, data-editing accounts, non-response analysis, guidelines for interviewer- and supervisors, maps, etc.

The general procedure for larger NIDI projects is that one researcher in a project-team is assigned to (also) act as data-manager of the project. His or her task is to organize how relevant project-related data and meta-data, during and after completion of a research project, are stored on the NIDI server (G: or Q: drives). This person also ensures that project-staff is well-informed about progress in database design, data editing and cleaning, so that, in principle, a colleague can take over the job during long-term absence of the data manager of the project. In principle, this data manager also has the duty to archive relevant project related data and meta-data at DANS. Before archiving, project staff members/project leader make final decisions about which data and meta-data must be stored at DANS, and what the access restrictions are, if applicable.

As NIDI projects are often carried out in collaboration with other, often foreign, institutes, data archiving and third-party access restrictions should be discussed with partners (and funding agencies) already at the design stage of projects. Therefore, a data archiving and third-party data-access paragraph should be included in project proposals. This implies that project proposals must include a distinct budget-line for all work involved in the archiving of project data and meta-data.

As NIDI research projects differ in complexity, it is not possible to list in this guideline which type of data and meta-data should be archived. The aforementioned criteria of *verifiability* and *reproducibility* will be of help in deciding what to archive and what not. In practice this means that any third-party team of monitoring and evaluation experts must have access to information and data necessary to replicate and verify all analytical work produced by the project. For a multi-country survey project this should be:

- General and country-specific questionnaires and interviewer guidelines
- Documents describing general and country-specific research designs, sample designs and sampling implementation accounts, fieldwork monitoring and evaluation accounts

- Documents describing data base design, and major data editing and data correction activities
- Data-description documents (e.g. variable and value labels) and documents that relates each data-item in the dataset to a specific item in the questionnaire. Only if there are particular analytical reasons to do so, derived or secondary variables could be *excluded* from data-files for archiving.
- Cleaned country-specific data set, (1), whereby all variables have clearly described variable and value labels, including ex-post-derived value-labels related to questionnaire answer-options “other, specify:.....”
- Cleaned country-specific data set (2), including un-standardized and standardized sample design weights (e.g. stratum weights, district-, neighbourhood-, household-, individual person weights).
- Documents describing results of fieldwork monitoring and evaluation experiences
- Final analytical project reports (general and country-specific)
- List of names and institutional affiliation of all project staff, and labelling of their main role in the project (e.g. project coordinator, data base advisor, interviewer supervisor, interviewer, project researcher, project financial staff-member, driver, secretary, etc.)
- Maps of areas where the research was implemented (as detailed as possible, without violating privacy protection of respondents).
- List of names and addresses of respondents (can be stored at DANS under the strictest embargo for reasons of privacy), permitting follow-up research under same respondents at a later date by the same research group or third-parties).

The proposed procedure at NIDI is that a project data manager, after completion of the project, first creates a distinct sub-subdirectory in the main directory Q:\DANS\_Easy\_Repository on the NIDI server. In this directory he or she copies all relevant data and meta-data for archiving, organized in sub-subdirectories. The data-manager creates one single zip-file comprising all files and subdirectories for archiving and uploads this file to DANS. After uploading at DANS this zip-file automatically unpacks and reproduces the subdirectory structure as specified by the NIDI data manager in the project’s subdirectory on the NIDI server (i.e. Q:\DANS\_Easy\_Repository\

The person responsible for archiving data and meta-data at DANS account starts out by opening an account on the DANS-EASY webpage (<https://easy.dans.knaw.nl/ui/home>). In the right-hand top corner of this webpage, click ‘Register’ to complete registration of a DANS-EASY account.



Get exposure and credit for your data: write a data paper for the new peer reviewed, online-only open access Research Data Journal (published by Brill)

For more info: [brill.com/rdj](http://brill.com/rdj)



## ARE YOUR RESEARCH DATA ARCHIVED SUSTAINABLY?

DEPOSIT YOUR DATA

The data manager fills in the forms and ensures that the *account is registered under his/her name*, including *personal E-mail address* to facilitate correspondence between prospective users of the data, DANS and the (NIDI) project data-manager who acts as account holder for the project data and meta data. Thus, the NIDI data-manager becomes *DANS accountholder* for a particular project.

The data-manager then proceeds with the archiving of project data and meta-data by clicking on ‘Deposit your data’ on the abovementioned webpage. On the subsequent web-page, select “Start deposit” under “Social and behavioural sciences”

### DEPOSIT YOUR DATA

To start depositing your data you will need to:

1. Prepare your data;
2. Log in to EASY or create an account;
3. Choose the relevant discipline below to start the deposit procedure.

More information about the deposit process can be found on the DANS website.

|   |  |
|---|--|
| <p>Archaeology</p> <p>Instructions in English or Nederlands (PDF).</p> <p>Start deposit</p>                     | <p>History</p> <p>Instructions in English or Nederlands (PDF).</p> <p>Start deposit</p>                    |
| <p>Social and behavioural sciences</p> <p>Instructions in English or Nederlands (PDF).</p> <p>Start deposit</p> | <p>Life sciences and medicine</p> <p>Instructions in English or Nederlands (PDF).</p> <p>Start deposit</p> |
| <p>Language and literature</p> <p>Instructions in English or Nederlands (PDF).</p> <p>Start deposit</p>         | <p>All other disciplines</p> <p>Instructions in English or Nederlands (PDF).</p> <p>Start deposit</p>      |

Before clicking “Start deposit” the *accountholder* should download and read the Instructions PDF as these will help him or her to fill in the ‘deposit’ forms and to upload the NIDI project data and meta-data. This “instruction PDF” as well as a document specifying “DANS preferred file-formats” are also present on the NIDI server (Q:\DANS\_Easy\_Repository). Depositing data at DANS should not take more than 30 minutes. The E-mail address of the accountholder will be the DANS E-mail address to which all correspondence about the project’s data and meta data will be sent. After clicking on “Start deposit” the DANS login screen is presented,

## LOG IN TO EASY

Please log in using your...

EASY ACCOUNT

Username

Password

[Log in](#)

[Can't access your account?](#)

[No EASY account yet?](#)

After login, 5 forms have to be completed before the zip-file of the project’s data and meta-data can be uploaded. In the first form, the name and other particulars of the CREATOR has to be entered. It is the name and affiliation entered in the fields belonging to the “Creator” that will be appear in all literature references. The name and affiliation of the creator is similar to the name of an “Author” of an article. So be aware which name is entered here.

## DEPOSIT DATASET - REQUIRED ELEMENTS

DEPOSIT:

1. Required elements
2. Upload files
3. Recommended elements
4. Additional elements
5. Overview and submitting

Instructions in [English](#) or [Nederlands](#) (PDF)

Creator 

|                         |                      |
|-------------------------|----------------------|
| (Academic) Title(s)     | <input type="text"/> |
| Initials                | <input type="text"/> |
| Prefix                  | <input type="text"/> |
| Surname                 | <input type="text"/> |
| Digital Author ID (DAI) | <input type="text"/> |
| Organization            | <input type="text"/> |



Person and/or organization responsible for the data

Title 



Title of the research project


Description 

Short description of the research project

The name and email-address entered in the “Creator” boxes may thus be different from the one of the Accountholder!. However, be aware that all correspondence regarding the projects data will arrive at the E-mail address of the account-holder, and not to the email-address of the creator!


The depositor of data and meta-data has to specify in form 1 whether or not particular access restrictions apply. In this way full control is guaranteed about who and when third-parties can download and use the deposited data and meta-data. The KNAW view is that restricted access may apply for a maximum of 18 months after the official date of project completion.

### Access

Access rights 

- Open access - Unrestricted access (CC0 Waiver No Rights Reserved - <https://creativecommons.org/about/cc0>)
- Open access for registered users - Unrestricted access for all registered EASY users
- Restricted: request permission - Registered EASY users, but only after depositor permission is granted
- Other access - The data are not available via EASY (they are either accessible in another way or elsewhere)

Description of the accessibility to the dataset files

Date available (optional) 


Date (yyyy-mm-dd) on which the dataset files will become accessible (future date acts as an embargo)

Audience 



Target group for the dataset

Save draft

Next »»

After the forms are completed and zip-file is uploaded and submitted (see form 5), the ACCOUNTHOLDER will receive a confirmation email from DANS stating that the archived files have been received and will be screened by DANS data managers for completeness and inconsistencies. Some correspondence may follow, possibly with the request to make certain adaptations in the files, and to re-upload (certain) the zip-file(s). The experience so far is that DANS data managers work efficient and quick.

At NIDI, the name and other particulars (E-mail address, DANS login-name and password) of the data manager who was responsible for archiving at DANS must be entered in a spreadsheet or Word document (overview for NIDI management) located in the main directory on the NIDI Q: drive (i.e. Q:\DANS\_Easy\_Repository). This is important because this NIDI data manager (i.e. DANS accountholder) may leave NIDI, so that a replacement person must be assigned and substituted as contact point for DANS. The incoming responsible must immediately alter the recorded ACCOUNTHOLDER name and email address at DANS, as this is the channel through which DANS communicates with NIDI about a project's data and meta-data.